

# Impact of loop statistics on the thermodynamics of RNA folding

Thomas R. Einert,<sup>1</sup> Paul Näger,<sup>1</sup> Henri Orland,<sup>2</sup> and Roland R. Netz<sup>1,\*</sup>

<sup>1</sup>*Physik Department, Technische Universität München, 85748 Garching, Germany*

<sup>2</sup>*Institut de Physique Théorique, CEA Saclay, 91191 Gif-sur-Yvette Cedex, France*

(Dated: July 29, 2008)

Loops are abundant in native RNA structures and proliferate close to the unfolding transition. By including a statistical weight  $\sim \ell^{-c}$  for loops of length  $\ell$  in the recursion relation for the partition function, we show that the heat capacity depends sensitively on the presence and value of the exponent  $c$ , even for a short explicit tRNA sequence. For long homo-RNA we analytically calculate the critical temperature and critical exponents which exhibit a non-universal dependence on  $c$ .

PACS numbers: 87.15.A-, 87.15.-v, 87.14.gn, 05.70.Fh

Apart from its role as an information carrier, RNA has regulatory and catalytic abilities[1]. Since this RNA functionality is mostly determined by its three-dimensional conformation, the accurate prediction of RNA folding from the base sequence is a central issue[2]. To a fairly good approximation RNA folding can be separated into the formation of a secondary structure, completely determined by the enumeration of all base pairs present in a given sequence, and the tertiary structure formation which only operates on the already existing secondary structural elements[3]. This constitutes a major simplification compared to the protein folding problem. Since the folding of even short RNA molecules takes much longer than reachable with all-atomistic simulations including explicit solvent, the more modest goal of obtaining the most probable secondary structures based on experimentally derived base-pairing and base-stacking free energies has been pursued[4, 5].

Due to the high number of unpaired bases, loops are abundant in RNA even at low temperatures. Polymer theory predicts the configurational weight of a loop consisting of  $\ell$  bases to decay as  $\ell^{-c}$  where the exponent  $c$  is universal and depends on the number of strands emerging from the loop[6]. We formulate the RNA partition function including the proper weight of loops using the same exponent  $c$  for terminal, internal, as well as multi-loops[7, 8]. For a homo-RNA in the thermodynamic limit a folding transition is known to exist in the finite range  $2 < c < 2.479$ [9, 10, 11]. We analytically calculate the  $c$ -dependent critical exponents of that transition. Critical effects are quite small which explains why they are not manifest in numerical calculations[12]. On the other hand, the non-critical effects of varying  $c$  are pronounced, even for real finite-length RNA sequences. We numerically calculate the heat capacity of a yeast tRNA with 76 bases using experimentally determined base pairing and stacking free energies[13]. At low temperatures the most probable structure consists of a characteristic clover-leaf structure and thus includes a multi-loop with four helices. Neglecting the loop statistics shifts the maximum of the heat capacity by more than 20 K, whereas including a realistic exponent  $c$  gives heat capacity curves that

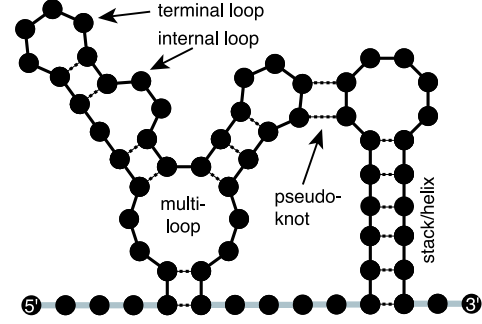


Figure 1: Schematic representation of a secondary RNA structure. Solid lines denote the RNA backbone, broken lines base pairs, and gray lines non-nested backbone bonds that are counted by the variable  $M$ ; here  $M = 11$ .

agree much better in shape with experimental results[14].

A primary RNA structure is fully determined by the base sequence  $\{b_N\}$  which is a list of nucleotides,  $b_i = \text{C, G, A or U}$  with  $N$  entries. In agreement with previous treatments, a valid secondary structure is a list of all base pairs with the constraint that a base can be part of at most one pair. In addition, pseudo-knots are not allowed, *i. e.* for any two base pairs  $(i, j)$  and  $(k, l)$  with  $i < j$ ,  $k < l$ , and  $i < k$  we have either  $i < k < l < j$  or  $i < j < k < l$ [15]. The statistical weight of a secondary structure depends on the free energy of base pair formation but also on the stacking energy of neighboring base pairs. For two neighboring pairs  $(i, j)$  and  $(i+1, j-1)$ , the free energy containing both pairing and stacking is  $g[(b_i, b_j), (b_{i+1}, b_{j-1})]$ . The statistical weight of a helical section starting with  $(i, j)$  and ending with  $(i+h, j-h)$  is  $w_{(i+h, j-h)}^{(i, j)} = \exp[-\beta(g^i[b_i, b_j] + \sum_{h'=1}^h g[(b_{i+h'-1}, b_{j-h'+1}), (b_{i+h'}, b_{j-h'})] + g^t[b_{i+h}, b_{j-h}])]$  where  $\beta^{-1} = k_B T$ . Here  $g^i$ ,  $g^t$  are initialization and termination free energies for base pairs located at the helix ends. All values  $g$ ,  $g^i$ ,  $g^t$  are extracted from experiments[13], see supplementary information[16].

In our notation, the canonical partition function  $Q_{i,j}^M$  of a sub-strand from base  $i$  at the 5' end through  $j$  at the 3' end depends on the number of non-nested backbone

bonds  $M$  [7, 10], see Fig. 1. The recursion relations for  $Q_{i,j}^M$  can be written as

$$Q_{i,j+1}^{M+1} = \frac{v_f(M+1)}{v_f(M)} \left[ Q_{i,j}^M + \sum_{k=i+M+1}^{j-N_{\text{loop}}} Q_{i,k-1}^M Q_{k,j+1}^0 \right] \quad (1a)$$

and

$$Q_{k,j+1}^0 = \sum_{h=1}^{(j-k-N_{\text{loop}})/2} w_{(k+h,j+1-h)}^{(k,j+1)} \times \sum_{m=1}^{j-k-1-2h} Q_{k+1+h,j-h}^m \frac{v_1(m+2)}{v_f(m)}. \quad (1b)$$

Eq. (1a) describes elongation of an RNA structure by either adding an unpaired base (first term) or by adding an arbitrary sub-strand  $Q_{k,j+1}^0$  that is terminated by a helix. Eq. (1b) constructs  $Q_{k,j+1}^0$  by closing structures with  $m$  non-nested bonds, summed up in  $Q_{k+1+h,j-h}^m$ , by a helix of length  $h$ .  $N_{\text{loop}} = 3$  is the minimum number of bases in a terminal loop.  $v_f(M)$  and  $v_1(M)$  denote the numbers of configurations of a free and a looped chain with  $M$  links, respectively, for which we use the asymptotic forms  $v_f(M) = y^M M^{\gamma-1}$  and  $v_1(M) = y^M M^{-c}$  [17]. The dependence on the monomer fugacity  $y$  and the exponent  $\gamma$  drops out by introducing the rescaled partition function  $\tilde{Q}_{i,j}^M = Q_{i,j}^M / (y^{j-i} M^{\gamma-1})$  and will not be considered further. The unrestricted partition function of the entire RNA is given by  $Z_N = \sum_M Q_{0,N}^M$ .

The loop exponent is  $c_{\text{ideal}} = 3/2$  for an ideal polymer and  $c_{\text{SAW}} = d\nu \simeq 1.76$  for an isolated self avoiding loop with  $\nu \simeq 0.588$  in  $d = 3$  dimensions [17]. However, helices which emerge from the loop increase  $c$  even further. In the asymptotic limit of long helical sections renormalization group predicts  $c_l = d\nu + \sigma_l - l\sigma_3$  for a loop with  $l$  emerging helices [6, 18] where  $\sigma_l = \epsilon l(2-l)/16 + \epsilon^2 l(l-2)(8l-21)/512 + \mathcal{O}(\epsilon^3)$  in an  $\epsilon = 4-d$  expansion. One obtains  $c_1 = 2.06$  for terminal,  $c_2 = 2.14$  for internal loops and  $c_4 = 2.16$  for a loop with four emerging helices. The variation of  $c$  over loop topologies that appear in the native structure of yeast tRNA-phe (shown in the inset Fig. 2) is thus quite small which justifies our usage of a constant exponent  $c$  for loops of all topologies. For larger  $l$  the  $\epsilon$  expansion prediction for  $c_l$  becomes unreliable. We therefore treat  $c$  as a heuristic input parameter which can be thought to account for other loop-length dependent effects (such as salt-dependent electrostatic loop self energies) as well.

We implement the recursion relation, Eq. (1), numerically using a free energy parameter set [13, 16] that allows for the wobble base pair GU in addition to the usual Watson-Crick pairs (GC and AU). The boundary conditions are  $Q_{i,j}^M = 0$  for  $M > j-i$ ,  $M < 0$ , or  $j-i < 0$ , except for the initial condition  $Q_{i,i-1}^{-1} = 1$ . In Fig. 2 we show the experimental heat capacity of the tRNA-phe of

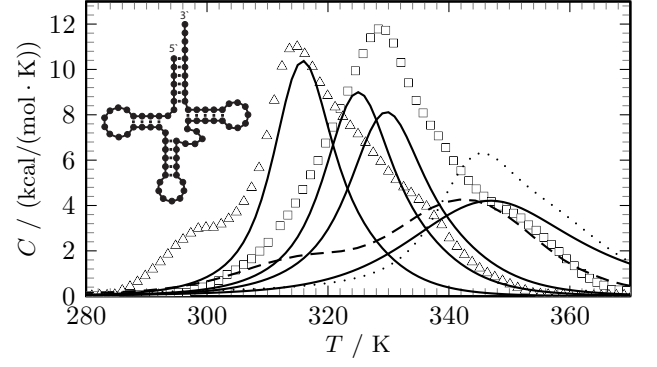


Figure 2: Experimental heat capacity of the tRNA-phe of yeast for NaCl concentrations 20 mM (triangles) and 150 mM (squares) [14]. Solid lines show results using Eq. (1) with loop exponents  $c = 3.0, 2.16, 1.76, 0$  (from left to right), compared with the results from the program **RNAheat** in the **Vienna package** [20] which uses a linearized multi-loop entropy for large loops (dashed curve). The dotted curve is obtained with  $c = 3$  and the same energy parameter set as for the solid curves, except for the loop initiation penalty which was omitted by setting  $g^i = g^t$ . The inset sketches the low-temperature secondary RNA structure obtained from Eq. (1), which perfectly matches experimental crystal structures.

yeast compared with our predictions from Eq. (1) using  $C = T\partial^2(k_B T \ln Z_N)/\partial T^2$ . The heat capacity peak corresponds to the gradual melting of the secondary structure. Although the RNA consists of just 76 nucleotides and is therefore far from the thermodynamic limit where one expects asymptotic effects to be important, the loop exponent  $c$  has drastic effects. Increasing  $c$  from 0 to 3 changes the peak width and height and decreases the melting temperature by more than 30 K (solid lines). In similar studies on DNA, where loops only appear close to the denaturation transition, the loop exponent was found to have much less influence [19]. In contrast, in RNA structures a large fraction of bases are unpaired and the correct modeling of loops is more important. It is difficult to directly compare experimental and theoretical curves, as the standard energy parameters used for secondary RNA-structure predictions are determined at 1 M NaCl concentration [13], while experimental heat capacity data is only available at 20 mM and 150 mM. Although the actual value of the loop exponent is not crucial (compare  $c = 1.76$  and  $c = 2.16$ ), the effect of neglecting loop statistics (i.e. setting  $c = 0$ ) is almost as big as omitting the loop initiation penalty contained in  $g^i$  (dotted line), a well established parameter, or changing the experimental salt concentration. Current secondary structure prediction tools approximate the entropy for large multi-loops by an affine function  $\ln(y^M M^{-c}) \approx \delta_0 + \delta_1 M$  [5, 20], which corresponds to a loop exponent  $c = 0$ . This is corroborated by the near agreement of the results from the Vienna package [20] (broken line) with the results from Eq. (1) using  $c = 0$ . It therefore is important to treat

the statistical weight of multi-loops on the same footing as terminal or internal loops, if they appear in the RNA groundstate, as is the case for tRNA. Our formulation of the RNA partition function can be generalized to more complicated loop weight functions to model the effects of salt or ligand binding.

We now consider homo-RNA, which can be realized experimentally with synthetic alternating sequences like  $[AU]_N$  or  $[GC]_N$ . The goal is to extract the critical asymptotic behavior embodied in Eq. (1) in the thermodynamic limit. We neglect base stacking, helix initiation and termination and simply give a statistical weight  $w = \exp[-\varepsilon/(k_B T)]$  to each base pair. This can be viewed as a coarse-graining approximation for natural or random RNA above the glass transition. Due to translational invariance Eqs. (1a) and (1b) simplify and can be combined to

$$\tilde{Q}_{N+1}^{M+1} = \tilde{Q}_N^M + w \sum_{n=M}^{N-1} \sum_{m=-1}^{N-n-2} \frac{\tilde{Q}_n^M \cdot \tilde{Q}_{N-n-2}^m}{(m+2)^c}, \quad (2)$$

where we introduced the total number of backbone segments  $N = j - i$  of a part which ranges from base number  $i$  through  $j$ . Next, we introduce the generating function

$$\mathcal{Z}(z, s) = \sum_{N=0}^{\infty} z^N \tilde{Z}_N(s) = \sum_{N=0}^{\infty} \sum_{M=0}^{\infty} z^N s^M \tilde{Q}_N^M. \quad (3)$$

For RNA with no external force one has  $s = 1$  and the sum over  $M$ , the number of non-nested backbone bonds, is unrestricted. In general,  $s = \exp(-\beta G(F)) > 1$  where  $G(F)$  is the change in free energy of a non-nested bond caused by a mechanical force  $F$  applied at the RNA ends. For a freely jointed chain one has  $G(F) = \beta^{-1} \ln[(\beta F a)^{-1} \sinh(\beta F a)]$ , with  $a$  being the Kuhn length. Combining Eqs. (2) and (3) yields

$$\mathcal{Z}(z, s) = \frac{\kappa(z)}{1 - s z \kappa(z)}. \quad (4)$$

where the function  $\kappa(z) = \sum_{N=0}^{\infty} z^N \tilde{Q}_N^0$  is the grand canonical partition function of an RNA with paired terminal bases. From Eqs. (3) and (4)  $\kappa(z)$  follows as the positive root of

$$\kappa(z)(\kappa(z) - 1) = w \text{Li}_c(z \kappa(z)), \quad (5)$$

where we use the polylogarithm  $\text{Li}_c(x) = \sum_{n=1}^{\infty} x^n / n^c$ . The thermodynamic behavior for  $N \rightarrow \infty$  is determined by the singularity of the generating function  $\mathcal{Z}(z, s)$  that is nearest to the origin in the complex  $z$ -plane. In particular, if  $z^*$  is the dominant singularity with  $\mathcal{Z}(z, s) \sim C_1(z^* - z)^\omega$  for  $z \rightarrow z^*$  and non-integer  $\omega$ , then  $\tilde{Z}_N(s) \sim C_1^{-1} N^{-(\omega+1)} z^{*- (N+1)}$  and the Gibbs free energy becomes to leading order  $\mathcal{G} = -k_B T \ln \tilde{Z}_N = N k_B T \ln z^*$ .

It turns out that  $\mathcal{Z}(z, s)$  has two singularities, first a branch point of  $\kappa(z)$  that follows by differentiating

Eq. (5) and whose position  $z_b$  is determined by

$$\kappa(z_b)^2 = w \text{Li}_{c-1}(z_b \kappa(z_b)) - w \text{Li}_c(z_b \kappa(z_b)). \quad (6)$$

Second, a simple pole  $z_p$  that follows from Eq. (4) and is determined by

$$s z_p \kappa(z_p) = 1. \quad (7)$$

The crossing of both singularities defines a critical point which is obtained by solving Eqs. (5)–(7) simultaneously. The critical base pairing weight  $w_{\text{cr}}$  as a function of the applied force fugacity  $s$  reads in closed form

$$w_{\text{cr}} = \frac{\text{Li}_{c-1}(s^{-1}) - \text{Li}_c(s^{-1})}{(\text{Li}_{c-1}(s^{-1}) - 2\text{Li}_c(s^{-1}))^2}. \quad (8)$$

In Fig. 3a we show the phase diagram of RNA in terms of  $w$  and  $s$  for different values of the loop exponent  $c$ .

Let us now consider the force-free case, *i.e.*  $s = 1$ . Eq. (8) simplifies to  $w_{\text{cr}} = (\zeta_{c-1} - \zeta_c)(\zeta_{c-1} - 2\zeta_c)^{-2}$  where  $\zeta_c = \text{Li}_c(1)$  is the Riemann zeta function. It immediately follows that  $w_{\text{cr}}$  is finite and non-zero only in the exponent range  $2 < c < c^*$  with  $c^* \simeq 2.479$  determined by  $\zeta_{c^*-1} - 2\zeta_{c^*} = 0$ . For  $c \rightarrow 2$  from above, Eq. (8) predicts  $w_{\text{cr}} \rightarrow 0$ . Thus, for  $c < 2$ , the RNA is always in the folded state and  $\tilde{Z}_N(s) \sim N^{-3/2} z_b^{-N}$  is characterized by the branch point  $z_b$  irrespective of how weak the pairing energy is [9, 11]. For  $c > c^*$  the RNA is always unfolded and  $\tilde{Z}_N(s) \sim z_p^{-N}$  and is determined by the simple pole. Right at the critical point, *i.e.* for  $2 < c < c^*$ ,  $z_{\text{cr}} = z_b = z_p$  and  $w = w_{\text{cr}}$ , the loop statistics become crucial and we obtain the new scaling

$$\tilde{Z}_N(s) \sim N^{(2-c)/(c-1)} z_{\text{cr}}^{-(N+1)}. \quad (9)$$

This gives rise to non-universal critical behavior. The specific heat possesses a weak non-analyticity at the  $(n+2)$ -order critical point, meaning that the  $n^{\text{th}}$  derivative with respect to temperature diverges as

$$\frac{d^n}{dT^n} C \sim |T_{\text{cr}} - T|^{-\chi}, \quad (10)$$

with  $\chi = n - (3 - c)/(c - 2)$  for  $T < T_{\text{cr}}$  and  $\chi = 1$  for  $T > T_{\text{cr}}$  and  $n$  being the integer with  $(c - 2)^{-1} - 1 < n < (c - 2)^{-1}$ , see Fig. 3b and the supplementary information[16]. For  $c \rightarrow c^*$  we have  $n \rightarrow 2$ ; for  $c \rightarrow 2$  we find  $n \rightarrow \infty$ . The fraction of paired bases is obtained *via* differentiation  $\theta = \partial \ln \mathcal{Z} / (N \partial \ln w)$  and reads

$$\theta = \frac{2\text{Li}_c(z_b \kappa(z_b))}{\text{Li}_{c-1}(z_b \kappa(z_b))}, \quad \text{for } T < T_{\text{cr}} \text{ and} \quad (11)$$

$$\theta = 1 - (1 + 4w\zeta_c)^{-1/2}, \quad \text{for } T > T_{\text{cr}}. \quad (12)$$

As before the singularity at the critical point is very weak and the  $n^{\text{th}}$  derivative of  $\theta$  exhibits a cusp

$$\frac{d^n}{dT^n} \theta \sim |T - T_{\text{cr}}|^\lambda, \quad (13)$$

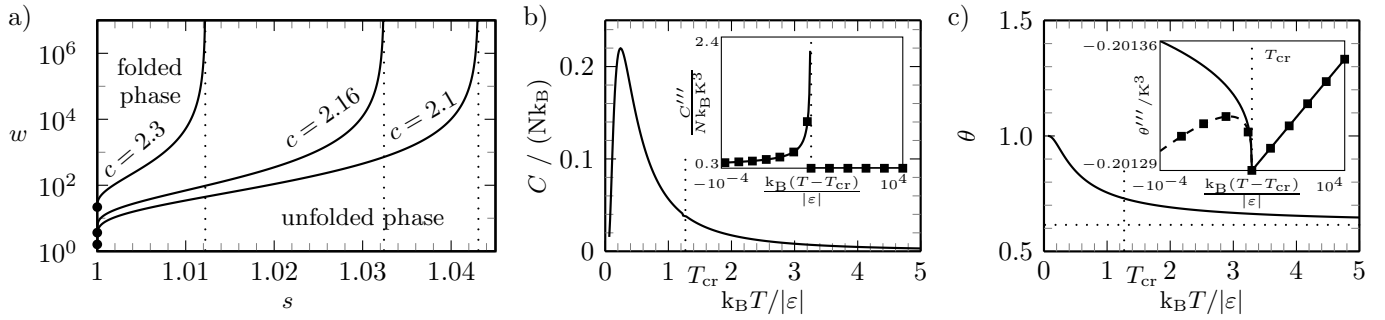


Figure 3: a) Phase diagram for three different values of the loop exponent  $c$  as a function of the base pairing weight  $w$  and force fugacity  $s$  featuring an unfolded phase (bottom) and a folded compact phase (top), following Eq. (8). The phase boundaries diverge at the vertical dotted lines. For  $c = c^* \simeq 2.479$  the phase boundary approaches  $s = 1$  and therefore only the unfolded phase exists. For  $c \leq 2$  there is only the folded phase. The dots denote the unfolding transition in the absence of external force, *i. e.*  $s = 1$ , which is considered in b) and c): Temperature dependence of the b) specific heat  $C$  and c) fraction of bound bases  $\theta$  for  $c = 2.3$ . The insets show the third derivatives  $C''' = d^3C/dT^3$  and  $\theta''' = d^3\theta/dT^3$  which clearly exhibit singular behavior. Squares denote numerical solutions of Eqs. (5) and (6) ( $T < T_{cr}$ ) or (7) ( $T > T_{cr}$ ). The solid lines show the leading order expansion around  $T_{cr}$  (denoted by vertical dotted lines), according to which  $C'''$  diverges with the exponent  $\chi = 2/3$  for  $c = 2.3$ , see Eq. (10), and  $\theta'''$  is characterized by the exponent  $\lambda = 1/3$ , see Eq. (13). In the inset of c) the analytical result including the next-leading order is also shown (broken line). The horizontal broken line in c) denotes the residual fraction of bound pairs at infinite temperature.

with  $\lambda = (c-2)^{-1} - n$  for  $T < T_{cr}$  and  $\lambda = 1$  for  $T > T_{cr}$ , see Fig. 3c where these asymptotic results are compared with numerical solutions of Eqs. (5)-(7).

We account for the asymptotic statistics of loops by including a loop-length dependent weight  $\ell^{-c}$  in the recursion relation for the partition function of RNA secondary structures. As a function of the loop exponent  $c$  we obtain exact critical exponents and boundaries between folded and unfolded phases for a simplified homopolymer model. Because the folding transition is at least of fourth order, the singular contribution to thermodynamic observables such as heat capacity or fraction of paired bases turns out to be quite small. On the other hand, the non-singular contribution at temperatures well below criticality depends crucially on  $c$ . This is demonstrated for an explicit sequence of a yeast tRNA by calculating heat capacities for various values of  $c$  and comparing with experimental data. It is seen that including realistic values for  $c$  is important and produces effects that are comparable to changing base pair stacking parameters or changing the salt concentration. So the conclusion is that while the dependence of critical properties on the loop exponent  $c$  is experimentally difficult to access and therefore largely irrelevant, the dependence of non-critical properties on  $c$  is important.

We are currently expanding the theory to allow for loop exponents that depend on the actual number of helices emerging from a given loop and to include tertiary contacts such as pseudo-knots or base triples which have been shown to play an important role in RNA folding[12].

Support from the *Elitenetzwerk Bayern* within the framework of *CompInt* is acknowledged.

- [1] R. F. Gesteland, T. R. Cech, and J. F. Atkins (eds.), *The RNA World* (Cold Spring Harbor Laboratory Press, Woodbury, 2005), 3<sup>rd</sup> edition.
- [2] I. Tinoco, O. C. Uhlenbeck, and M. D. Levine, *Nature* **230**, 362 (1971).
- [3] D. Thirumalai, *Proc. Nat. Acad. Sci.* **95**, 11506 (1998).
- [4] M. Zuker and P. Stiegler, *Nucleic Acids Res.* **9**, 133 (1981).
- [5] J. S. McCaskill, *Biopolymers* **29**, 1105 (1990).
- [6] B. Duplantier, *Phys. Rev. Lett.* **57**, 941 (1986).
- [7] R. Bundschuh and U. Gerland, *Phys. Rev. Lett.* **95**, 208104 (2005).
- [8] U. Gerland, R. Bundschuh, and T. Hwa, *Biophys. J.* **81**, 1324 (2001).
- [9] P.-G. de Gennes, *Biopolymers* **6**, 715 (1968).
- [10] M. Müller, F. Krzakala, and M. Mézard, *Eur. Phys. J. E* **9**, 67 (2002).
- [11] M. Müller, *Phys. Rev. E* **67**, 021914 (2003).
- [12] M. Baiesi, E. Orlandini, and A. L. Stella, *Phys. Rev. Lett.* **91**, 198102 (2003).
- [13] D. H. Mathews *et al.*, *J. Mol. Biol.* **288**, 911 (1999).
- [14] P. L. Privalov and V. V. Filimonov, *J. Mol. Biol.* **122**, 447 (1978).
- [15] H. Orland and A. Zee, *Nucl. Phys. B* **620**, 452, (2002).
- [16] See EPAPS document No. [...] for expansion coefficients and energy parameters. For more information on EPAPS, see <http://www.aip.org/pubservs/epaps.html>.
- [17] P.-G. de Gennes, *Scaling Concepts in Polymer Physics* (Cornell University Press, Ithaca, 1979).
- [18] Y. Kafri, D. Mukamel, and L. Peliti, *Phys. Rev. Lett.* **85**, 4988 (2000).
- [19] R. Blossey and E. Carlon, *Phys. Rev. E* **68**, 061911 (2003).
- [20] I. L. Hofacker *et al.*, *Mon. Chem.* **125**, 167 (1994).

\* E-mail: netz@ph.tum.de

# Supplementary material for: “Impact of loop statistics on the thermodynamics of RNA folding”

Thomas R. Einert,<sup>1</sup> Paul Näger,<sup>1</sup> Henri Orland,<sup>2</sup> and Roland R. Netz<sup>1,\*</sup>

<sup>1</sup>*Physik Department, Technische Universität München - 85748 Garching, Germany*

<sup>2</sup>*Institut de Physique Théorique, CEA Saclay, 91191 Gif-sur-Yvette Cedex, France*

(Dated: July 29, 2008)

We give the coefficients of the expansion of the branch point position  $z_b$  and the value  $\kappa_b$  of the function  $\kappa(z)$  in the vicinity of the critical point. We also list the free energy parameters used in our numerical study of the melting curve of the yeast tRNA-phe.

## EXPANSION OF THE BRANCH POINT NEAR THE CRITICAL POINT

For  $T < T_{\text{cr}}$  the dominant singularity of  $\mathcal{Z}(z)$ , which determines the thermodynamics, is the branch point at  $z = z_b$ . The branch point has its origin in the singular behavior of the function of  $\kappa(z)$ , which is determined by the equation

$$\kappa(z)(\kappa(z) - 1) = w \text{Li}_c(z\kappa(z)) . \quad (1)$$

In this section we are considering the asymptotic behavior of the branch point position  $z_b(T)$  and the functional value  $\kappa_b(T) := \kappa(z_b(T))$  at the branch point in the vicinity of the phase transition, in particular for  $T \rightarrow T_{\text{cr}}$  while  $T < T_{\text{cr}}$ . The existence of the phase transition implies for the loop exponent  $2 < c < c^* \approx 2.479$ .

The equation determining the position  $z_b$  of the branch point follows by differentiating Eq. (1) with respect to  $\kappa$ , leading to

$$\kappa(z_b)^2 = w \text{Li}_{c-1}(z_b \kappa(z_b)) - w \text{Li}_c(z_b \kappa(z_b)) \quad (2)$$

where  $d \text{Li}_c(x)/dx = \text{Li}_{c-1}(x)/x$  is used. For  $T > T_{\text{cr}}$  the pole  $z_p$  of  $\mathcal{Z}(z)$  is the dominant singularity. It follows from the equation

$$z_p \kappa(z_p) = 1 . \quad (3)$$

For the calculation of critical exponents, we always consider the force-free case characterized by  $s = 1$  where the RNA is not stretched externally.

## Scaling of the partition function at $T = T_{\text{cr}}$

The order of the branch point exactly at  $T = T_{\text{cr}}$  is calculated by expanding Eq. (1) in powers of  $z/z_{\text{cr}} - 1$  and  $\kappa(z)/\kappa_{\text{cr}} - 1$  while keeping  $w = w_{\text{cr}}$  fixed. Note that both Eq. (2) and (3) hold at  $T = T_{\text{cr}}$ . One obtains

$$\frac{\kappa(z) - \kappa_{\text{cr}}}{\kappa_{\text{cr}}} \sim - \left( - \frac{z - z_{\text{cr}}}{z_{\text{cr}}} \frac{\zeta_{c-1}}{\Gamma(1-c)} \right)^{\frac{1}{c-1}} . \quad (4)$$

Thus, the asymptotic behavior of the generating function at  $T = T_{\text{cr}}$  is

$$\mathcal{Z}(z, s = 1) = \frac{\kappa(z)}{1 - z\kappa(z)} \sim \kappa_{\text{cr}} \left( - \frac{z - z_{\text{cr}}}{z_{\text{cr}}} \frac{\zeta_{c-1}}{\Gamma(1-c)} \right)^{-\frac{1}{c-1}} . \quad (5)$$

$\mathcal{Z}(z, s) \sim C_1(z - z_{\text{cr}})^\omega$ , for  $z \rightarrow z_{\text{cr}}$  and non-integer  $\omega$ , implies  $Z_N \sim C_1^{-1} N^{-(\omega+1)} z_{\text{cr}}^{-(N+1)}$  and one obtains

$$Z_N \sim N^{(2-c)/(c-1)} z_{\text{cr}}^{-(N+1)} . \quad (6)$$

### Expansion of the branch point for $T < T_{\text{cr}}$

To obtain the critical behavior for  $T < T_{\text{cr}}$  we perform an asymptotic expansion of Eqs. (1) and (2) around the critical point, where the branch point and the pole coincide. Thus, at the critical point all Eqs. (1-3) have to hold and we obtain the critical values exactly as

$$\kappa_{\text{cr}} = \frac{1}{2} \left( 1 + \sqrt{1 + 4w\zeta_c} \right), \quad z_{\text{cr}} = 2 \left( 1 + \sqrt{1 + 4w\zeta_c} \right)^{-1}, \quad \text{and} \quad w_{\text{cr}} = \frac{\zeta_{c-1} - \zeta_c}{(\zeta_{c-1} - 2\zeta_c)^2}. \quad (7)$$

As an ansatz for  $z_{\text{b}}(T)$  and  $\kappa_{\text{b}}(T)$  we use a power series in  $d = w/w_{\text{cr}} - 1$ , where  $w = \exp[-\varepsilon/(k_{\text{B}}T)]$  is the Boltzmann weight of a base pair. To simplify notation, we define  $\alpha = (c - 2)^{-1}$  and write

$$z_{\text{b}}/z_{\text{cr}} \sim 1 + a_1 d + a_2 d^2 + \dots + d^\alpha (a_\alpha + a_{\alpha+1} d + a_{\alpha+2} d^2 + \dots) + a_{2\alpha-1} d^{2\alpha-1} + \dots \quad (8a)$$

$$\kappa_{\text{b}}/\kappa_{\text{cr}} \sim 1 + b_1 d + b_2 d^2 + \dots + d^\alpha (b_\alpha + b_{\alpha+1} d + b_{\alpha+2} d^2 + \dots) + b_{2\alpha-1} d^{2\alpha-1} + \dots \quad (8b)$$

Plugging the ansatz (8) into Eqs. (1) and (2) we can solve order by order. To do so, the series representation of the polylogarithm  $\text{Li}_\nu(x)$  around  $x = 1$  is used [2]

$$\text{Li}_\nu(x) \sim \zeta_\nu - \zeta_{\nu-1}(1-x) + \frac{1}{2}(\zeta_{\nu-2} - \zeta_{\nu-1})(x-1)^2 + \dots (1-x)^{\nu-1} \left( \Gamma(1-\nu) + \frac{1}{2}(1-\nu)\Gamma(1-\nu)(1-x) + \dots \right). \quad (9)$$

We obtain the coefficients for  $z_{\text{b}}$

$$a_1 = -\frac{\zeta_c}{\zeta_{c-1}} \quad (10a)$$

$$a_2 = \frac{\zeta_c^2}{\zeta_{c-1}^3} (2\zeta_{c-1} - \zeta_c) \quad (10b)$$

$$a_3 = -\frac{\zeta_c^3}{\zeta_{c-1}^5} (5\zeta_{c-1}^2 - 6\zeta_{c-1}\zeta_c + 2\zeta_c^2) \quad (10c)$$

$$a_4 = \frac{\zeta_c^4}{\zeta_{c-1}^7} (14\zeta_{c-1}^3 - 28\zeta_{c-1}^2\zeta_c + 20\zeta_{c-1}\zeta_c^2 - 5\zeta_c^3) \quad (10d)$$

$$a_5 = -\frac{\zeta_c^5}{\zeta_{c-1}^9} (42\zeta_{c-1}^4 - 120\zeta_{c-1}^3\zeta_c + 135\zeta_{c-1}^2\zeta_c^2 - 70\zeta_{c-1}\zeta_c^3 + 14\zeta_c^4) \quad (10e)$$

$$\vdots$$

$$a_\alpha = 0 \quad (10f)$$

$$a_{\alpha+1} = -\frac{\Gamma(1-c) + \Gamma(2-c)}{\zeta_{c-1}} \left( -\frac{\zeta_{c-1}^2 - 3\zeta_{c-1}\zeta_c + 2\zeta_c^2}{\Gamma(2-c)\zeta_{c-1}} \right)^{\alpha+1} \quad (10g)$$

$$a_{\alpha+2} = \frac{(\Gamma(1-c) + \Gamma(2-c))(\zeta_{c-1}^2 + 4(c-2)\zeta_{c-1}\zeta_c + (5-3c)\zeta_c^2)}{(c-2)\zeta_{c-1}^3} \left( -\frac{\zeta_{c-1}^2 - 3\zeta_{c-1}\zeta_c + 2\zeta_c^2}{\Gamma(2-c)\zeta_{c-1}} \right)^{\alpha+1} \quad (10h)$$

$$\vdots$$

$$a_{2\alpha-1} = 0 \quad (10i)$$

$$\vdots$$

We obtain the coefficients for  $\kappa_{\text{b}}$

$$b_1 = \frac{\zeta_c}{\zeta_{c-1}} \quad (11a)$$

$$b_2 = -\frac{\zeta_c^2}{\zeta_{c-1}^3} (\zeta_{c-1} - \zeta_c) \quad (11b)$$

$$b_3 = 2\frac{\zeta_c^3}{\zeta_{c-1}^5} (\zeta_{c-1} - \zeta_c)^2 \quad (11c)$$

$$b_4 = -5 \frac{\zeta_c^4}{\zeta_{c-1}^7} (\zeta_{c-1} - \zeta_c)^3 \quad (11d)$$

$$b_5 = 14 \frac{\zeta_c^5}{\zeta_{c-1}^9} (\zeta_{c-1} - \zeta_c)^4 \quad (11e)$$

$$\vdots$$

$$b_\alpha = - \left( - \frac{\zeta_{c-1}^2 - 3\zeta_{c-1}\zeta_c + 2\zeta_c^2}{\Gamma(2-c)\zeta_{c-1}} \right)^\alpha \quad (11f)$$

$$b_{\alpha+1} = \frac{\Gamma(1-c) + \Gamma(2-c)}{\zeta_{c-1}} \left( - \frac{\zeta_{c-1}^2 - 3\zeta_{c-1}\zeta_c + 2\zeta_c^2}{\Gamma(2-c)\zeta_{c-1}} \right)^{\alpha+1} + \left( - \frac{\zeta_{c-1}^2 - 3\zeta_{c-1}\zeta_c + 2\zeta_c^2}{\Gamma(2-c)\zeta_{c-1}} \right)^\alpha \frac{\zeta_{c-1}^2 - (c-2)\zeta_{c-1}\zeta_c - \zeta_c^2}{(c-2)\zeta_{c-1}^2} \quad (11g)$$

$$b_{\alpha+2} = \frac{\zeta_{c-1} - \zeta_c}{2(c-2)^2 \cdot \Gamma(2-c)\zeta_{c-1}^4} \left( - \frac{\zeta_{c-1}^2 - 3\zeta_{c-1}\zeta_c + 2\zeta_c^2}{\Gamma(2-c)\zeta_{c-1}} \right)^\alpha \times \left( 2(c-2)\Gamma(1-c)(\zeta_{c-1} - 2\zeta_c)(\zeta_{c-1}^2 + 2(c-2)\zeta_{c-1}\zeta_c + (5-3c)\zeta_c^2) + \Gamma(2-c)((c-3)\zeta_{c-1}^3 + (c-3)(4c-7)\zeta_{c-1}^2\zeta_c - (3c-5)(4c-9)\zeta_{c-1}\zeta_c^2 + (3c-5)(4c-7)\zeta_c^3) \right) \quad (11h)$$

$$\vdots$$

$$b_{2\alpha-1} = - \frac{\zeta_{c-2} - 3\zeta_{c-1} + 2\zeta_c}{(c-2)\Gamma(2-c)} \left( - \frac{\zeta_{c-1}^2 - 3\zeta_{c-1}\zeta_c + 2\zeta_c^2}{\Gamma(2-c)\zeta_{c-1}} \right)^{2\alpha-1} \quad (11i)$$

$$\vdots$$

Remarkably, the expansion of the product  $z_b \kappa_b$  yields

$$z_b \kappa_b \sim 1 + (a_\alpha + b_\alpha) d^\alpha + (a_{\alpha+1} + b_1 a_\alpha + b_{\alpha+1} + a_1 b_\alpha) d^{\alpha+1} + \mathcal{O}(d^{\alpha+2}), \quad (12)$$

meaning that all integer powers  $d^n$ , for  $n < \alpha$ , vanish. This fact renders the transition of high order when  $c \rightarrow 2$  and thus  $\alpha \rightarrow \infty$ . For the fraction of bound bases this can be seen directly by expanding the low temperature expression

$$\theta = 2 \frac{\text{Li}_c(z_b \kappa_b)}{\text{Li}_{c-1}(z_b \kappa_b)} \quad (13)$$

using Eqs. (9) and (12).

# ENERGY PARAMETERS

The enthalpy  $h$  and entropy  $s$  parameters used in our calculations are taken from reference [1]. For instance, the entries in the row UA and the column GC,  $h_{\text{UA,GC}}$  and  $s_{\text{UA,GC}}$ , give the enthalpy and entropy contribution due to the stacking of the two neighboring base pairs UA and GC, where U and C are located at the 5'-end. The bottom two rows contain the initiation and termination contribution. For instance, the total free enthalpy of the triple helix  $5'\text{-CGA-3'}$   $3'\text{-GCU-5'}$  is  $g_3 = (h_{\text{CG,GC}} + h_{\text{GC,AU}} + h_{\text{CG}}^i + h_{\text{AU}}^t) - T(s_{\text{CG,GC}} + s_{\text{GC,AU}} + s_{\text{CG}}^i + s_{\text{AU}}^t)$ .

	Enthalpy $h$ / (kcal/mol)							Entropy $s$ / ( $10^{-3}$ kcal/(mol K))					
	AU	UA	CG	GC	GU	UG		AU	UA	CG	GC	GU	UG
AU	-6.82	-9.38	-11.40	-10.48	-3.21	-8.81	AU	-19.0	-26.7	-29.5	-27.1	-8.6	-24.0
UA	-7.69	-6.82	-12.44	-10.44	-6.99	-12.83	UA	-20.5	-19.0	-32.5	-26.9	-19.3	-37.3
CG	-10.44	-10.48	-13.39	-10.64	-5.61	-12.11	CG	-26.9	-27.1	-32.7	-26.7	-13.5	-32.2
GC	-12.40	-11.40	-14.88	-13.39	-8.33	-12.59	GC	-32.5	-29.5	-36.9	-32.7	-21.9	-32.5
GU	-12.83	-8.81	-12.59	-12.11	-13.47	-14.59	GU	-37.3	-24.0	-32.5	-32.2	-44.9	-51.2
UG	-6.99	-3.21	-8.33	-5.61	-9.26	-13.47	UG	-19.3	-8.6	-21.9	-13.5	-30.8	-44.9
$h^i$	7.33	7.33	3.61	3.61	7.33	7.33	$s^i$	9.0	9.0	-1.5	-1.5	9.0	9.0
$h^t$	3.72	3.72	0.00	0.00	3.72	3.72	$s^t$	10.5	10.5	0.0	0.0	10.5	10.5

\* E-mail: netz@ph.tum.de

- [1] D. H. Mathews, J. Sabina, M. Zuker, and D. H. Turner, J. Mol. Biol. **288**, 911 (1999).  
[2] T. S. Erdélyi, *Higher Transcendental Functions 1* (McGraw-Hill, New York, 1953).